



EUDAT

Towards a pan-European Collaborative Data
Infrastructure

Damien Lecarpentier
CSC-IT Center for Science, Finland
DATA2012, Indianapolis



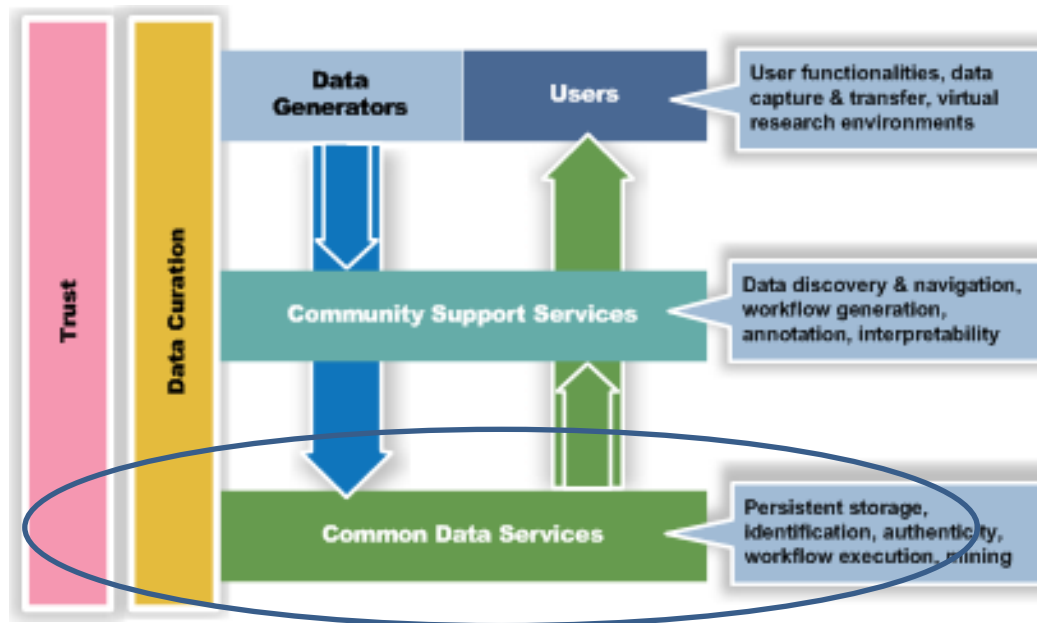
Outline of the talk

- ❑ EUDAT concept
- ❑ EUDAT consortium
- ❑ EUDAT service approach
- ❑ Some challenges ahead

EUDAT Key facts

Project Name	EUDAT – European Data
Start date	1st October 2011
Duration	36 months
Budget	16,3 M€ (including 9,3 M€ from the EC)
EC call	Call 9 (INFRA-2011-1.2.2): Data infrastructure for e-Science (11.2010)
Participants	25 partners from 13 countries (national data enters, technology providers, research communities, and funding agencies)
Objectives	“To deliver cost-efficient and high quality Collaborative Data Infrastructure (CDI) with the capacity and capability for meeting researchers’ needs in a flexible and sustainable way, across geographical and disciplinary boundaries.”

The CDI concept



EUDAT Core Service Areas

Community-oriented services

- Simple Data Access and upload
- Long term preservation
- Shared workspaces
- Execution and workflow (data mining, etc.)
- Joint metadata and data visibility

Enabling services (making use of existing services where possible)

- Persistent identifier service (EPIC, DataCite)
- Federated AAI service
- Network Services
- Monitoring and accounting

Core services are building blocks of EUDAT's Common Data Infrastructure
mainly included on bottom layer of data services



Consortium



Consortium



Research Communities

ESFRI

CLARIN
Common Language Resources and Technology Infrastructure



January 20, 2011
New Virtual Language Observatory launch

Semantic data description and descriptive metadata are vital factors for determining if the data can be reused in the future. These metadata are still dependent on rapidly changing ontologies and terminologies.

John Marks
ESF 2008

Activities
Publications
Solutions
Laboratory
Consultancy
Virtual Language Observatory

Internal Web Site

EPOS
EUROPEAN PLATE OBSERVING SYSTEM

Research Infrastructure and E-Science for Data and Observatories on Earthquakes, Volcanoes, Surface Dynamics and Tectonics



Mission & Vision Objectives Architecture Partners Preparatory Phase Data Products

LIFEWATCH

e-science and technology infrastructure for biodiversity data and observatories

Home Contact About Participants Get Involved News Cases Events Press Documents

LIFEWATCH COUNTRIES

Austria Belgium Denmark Finland France Greece Hungary Italy Netherlands Norway Poland
Portugal Romania Slovak Republic Slovenia Spain Sweden Turkey United Kingdom

LIFEWATCH NEWS

2011-02-16 LIFEWATCH RESEARCH INFRASTRUCTURE STARTS CONSTRUCTION IN 2011 - The initial country consortium establishing the Lifewatch research infrastructure agreed to finance ... [Read more](#)

2011-01-19 LIFEWATCH CLOSING EVENT - On this page you can download all the slides presented at the closing event of the Lifewatch preparatory project a first group of ... [Read more](#)

2011-01-17 LIFEWATCH CONSTRUCTION KICKS OFF ON JANUARY 19TH - On 19 January 2011, at the closing conference of the Lifewatch preparatory project a first group of ... [Read more](#)

LIFEWATCH FOCUS

Lifewatch research infrastructure starts construction in 2011
The initial country consortium establishing the Lifewatch research infrastructure agreed to finance the start-up activities for the infrastructure construction. These countries will host the Common Facilities of Lifewatch.
On 19th January 2011 representatives from organizations in Hungary, Italy, the Netherlands, Romania and Spain signed a Memorandum of Understanding to cooperate for an early start of the Lifewatch infrastructure for biodiversity and ecosystem research. The Lifewatch Stakeholders Board, representing the ten countries aiming at establishing the Lifewatch ERIC, welcomed the initiative to start early construction.

Newsletter
Subscribe to our newsletter. Send an email to newsletter@lifewatch.eu

Quote
"Through our Memorandum of Cooperation GBIF and Lifewatch based on our respective complementary mandates, now have a formal framework for co-operation and collaboration on infrastructural developments, building on GBIF's 10 years of investment to date."
Dr. Nick King
Director Global Biodiversity Information Facility (GBIF)

enes
European Network for Earth System Modelling

Welcome

News

IS-ENES

The Rationale

The Aims

The MOU

Related Projects

Partners

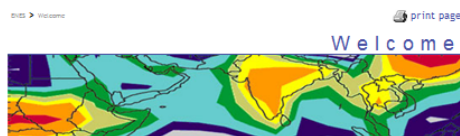
Meetings

Sitemap

Contact

Imprint

search...



ENES Townhall Meeting at EGU 2010: Here is the [announcement!](#)

For latest news on IS-ENES click [here!](#)

A major challenge for the climate research community is the development of comprehensive Earth system models capable of simulating natural climate variability and human-induced climate changes. Such models need to account for detailed processes occurring in the atmosphere, the ocean and on the continents including physical, chemical and biological processes on a variety of spatial and temporal scales. They have also to capture complex nonlinear interactions between the different components of the Earth system and assess, how these interactions can be perturbed as a result of human activities.

Accurate scientific information is required by government and industry to make appropriate decisions regarding our global environment, with direct consequences on the economy and lifestyles. It is therefore the responsibility of the scientific community to accelerate progress towards a better understanding of the processes governing the Earth system and towards the development of an improved predictive capability. An important task is to develop an advanced software and hardware environment in Europe, under which the most advanced high resolution climate models can be developed, improved, and integrated.



Project

Home
News
Objectives
VPH 2010
Calendar
VPH Lines
Contact Us
Archives

Activities

European Project
Resources
Publications
Dissemination material
VPH Events
Event Archive

Home WP1 WP2 WP3 WP4 WP5 VPH-I MIP Login

Welcome to the home page of the Virtual Physiological Human Network of Excellence (VPH NoE) and information portal for the VPH Initiative

VPH NoE

BioMed Town

VPH Initiative

VPH for the Public

VPH for Researchers

VPH for Clinicians

The VPH NoE is a project which aims to help support and progress European research in biomedical modelling and simulation of the human body. This will improve our ability to predict, diagnose and treat disease, and have a dramatic impact on the future of healthcare, the pharmaceutical and medical device industries.

VPH 2010 September 30th - October 1st 2010 Brussels, Belgium

SEVENTH FRAMEWORK PROGRAMME

Home WP1 WP2 WP3 WP4 WP5 VPH-I MIP Login

Building a wider VPH Community

HIGHLIGHTS

Interface Focus special issue with best papers from the VPH2010 Conference

VPH NoE and the Pistoia Alliance

Example Project Call 3:

Join the Public Forum of the VPH-IET Support Action

Multi-Institutional Graduate Programme for Virtual Physiological Human Sciences (VPH-I-MGP)

VPH Vision & Strategy Paper I

VPH NoE 2010 Newsletter (Jan 2010) now available

LATEST VPH EVENTS

- 01.06.2011 - 03.06.2011 ICOS 2011 (Toulon)
- 06.06.2011 - 09.06.2011 VPH08
- 08.06.2011 - 10.06.2011 VPH08
- 08.06.2011 - 10.06.2011 VPH08

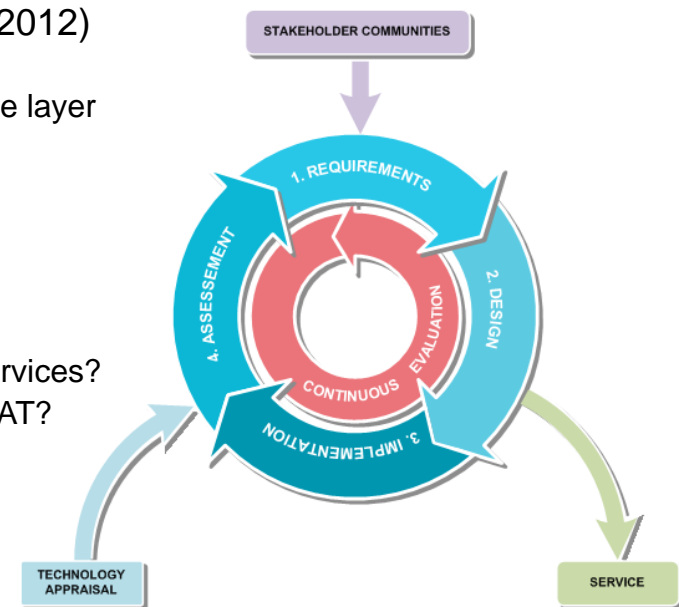
EUDAT service design activities

1. Capturing Communities Requirements (WP4)

- 1st round of interviews with the five initial communities (Oct.-Dec. 2012)
 - Understand how data is organised in each community
 - Collect first wishes and specific requirements from a common data service layer
- Next phase: refine analysis and expanding it to other communities

2. Building the corresponding services (WP5)

- Technology appraisal (ongoing)
 - What is already available at partners's sites to build the corresponding services?
 - What are the gaps and market failures that should be addressed by EUDAT?
- Next phase: Developing candidate services
 - Adapt services to match the requirements
 - Integrate with community and SP services
 - Test and evaluate with communities



3. Deploying the services and operating the federated infrastructure (WP6)

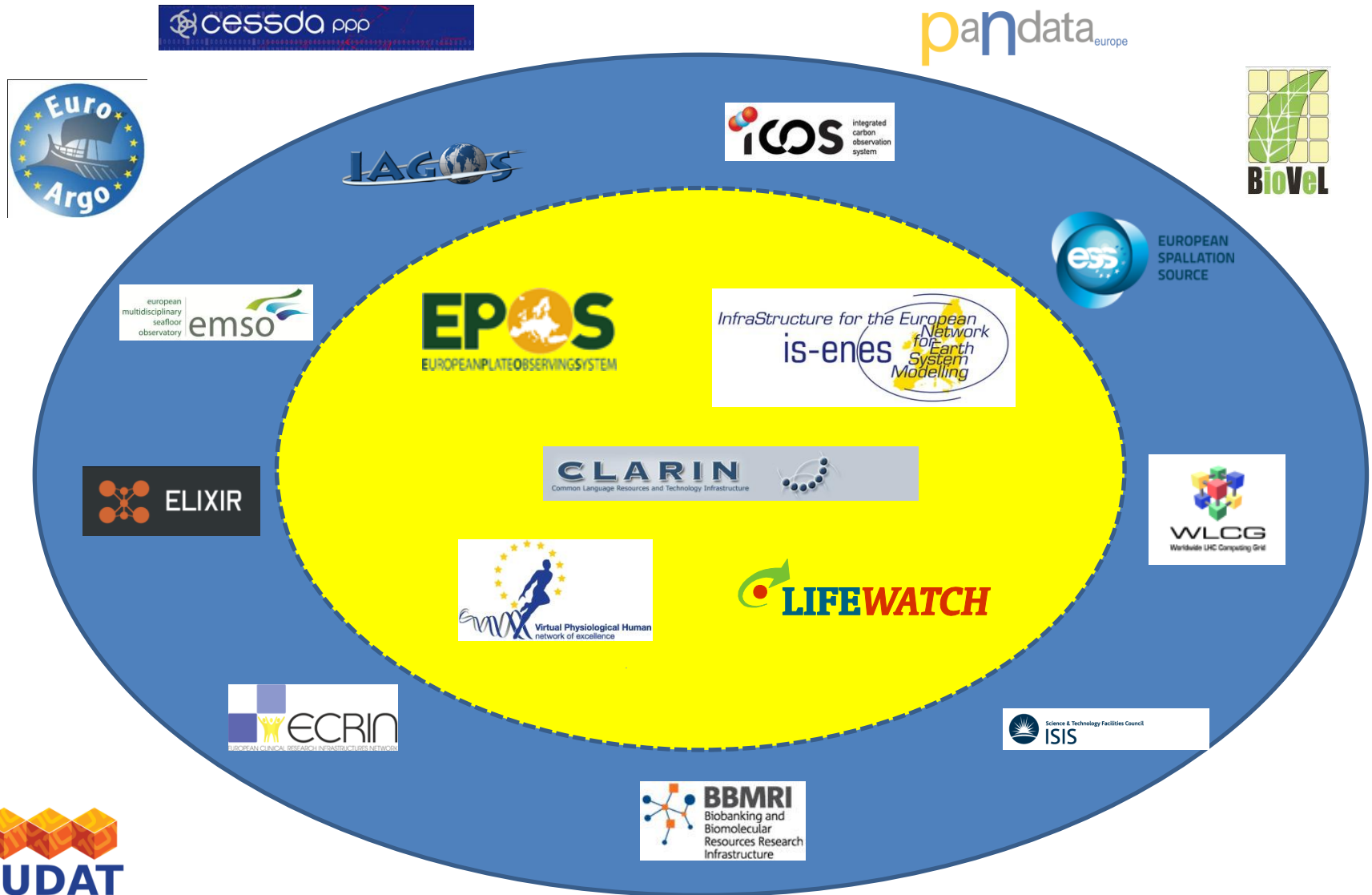
- Designing the federated infrastructure and the interfaces for cross-site operations (ongoing)
- Next phase: integrating and coordinating resource provision, operations and support

First Service Cases

■ November 2011: shortlist of 6 service/use cases identified

- **Safe replication:** Allow communities to replicate data to selected data centers for storage and do this in a robust, highly available way. **Under implementation**
- **Dynamic replication:** Perform (HPC) computations on replicated data. Move (part of) the safely replicated data to powerful machines and move the results back into the archives. **Under implementation**
- **AAI:** A solution for a working AAI system in a federation scenario.
- **Metadata:** A joint metadata domain for all data that is stored by EUDAT data centers by harvesting metadata records for all data objects from the communities. Allow to have a catalogue to demonstrate what EUDAT stores, and to have a registry which can be used for automatic operations such as data mining.
- **PID:** a robust, highly available and effective PID system that can be used within the communities and by EUDAT.
- **Research data store:** A function that will help researchers mediated by the participating communities to upload and store data which is not part of the officially handled data sets of the community.

Reaching out to other communities



Research fields

Environmental Science	ENES, EPOS, Lifewatch, EMSO, IAGOS-ERI, ICOS, Euro-Argo
Social Sciences and Humanities	CLARIN, CESSDA, DARIAH
Biological and Medical Science	VPH, ELIXIR, BBRMI, ECRIN, DiXA
Physical Sciences and Engineering	WLCG, ISIS, PanData
Material Science	ESS...

EUDAT targets all scientific disciplines (discipline neutral):

- To enable the capture and identify cross-discipline requirements
- To involving the scientists of all the communities in the shaping of the infrastructure and its services



Some forthcoming challenges

■ Scaling

- Can we expect that the requirements identified will be shared by other research communities?
- How to maintain a high level of interoperability in the context of diversity of data, disciplines and practices?
- How far shall the geographical pan-European dimension be sought for?

■ Collaboration models

- What kind of collaboration and partnership can we offer to interested stakeholders?
- How will the future infrastructure work with and interact with other infrastructures and projects?
- How to move for a project collaboration to a federated infrastructure?

■ Funding models

- How will the infrastructure be funded in the future?
- How far are the public bodies (EC and MS) willing to fund the infrastructure?
- Shall users pay for the service they are provided, and if so on what basis (pay per use, membership fee, etc.)?

EUDAT User Forum



EUDAT - 1ST USER FORUM
TOWARDS A COLLABORATIVE DATA INFRASTRUCTURE
INVESTIGATING RESEARCH COMMUNITIES REQUIREMENTS

Place:
Jordi Girona Street, No.31
Rectorat Building, Juntres room,
08034, Barcelona

DAY 1: 7TH MARCH 2012

12.00	Welcome and Snacks	
13.00	SESSION 1 - SETTING THE SCENE	CHAIR: KIMMO KOSKI
13.00	EUDAT - an Overview for a User Perspective	Damien Lecarpentier
13.15	CDI and EUDAT	Peter Wittenburg
13.30	Climate modeling and EUDAT	Michael Lautenschlager
13.45	Seismology and EUDAT	Alberto Michelini
14.00	Physiology and EUDAT	Stefan Zasada
14.15	Linguistics and EUDAT	Pavel Straňák/Daan Broeder
14.30	Discussion	
15.00	Coffee Break	
15.30	SESSION 2 - EUDAT SERVICE CASES	CHAIR: ALBERTO MICHELINI
15.30	Replication service and its requirements	Peter Wittenburg
15.40	Staging Replicas for computations and requirements	Stefan Zasada
15.50	Researchers' Data Store and requirements	Daan Broeder
16.00	Joint EUDAT Metadata Domain	Michael Lautenschlager
16.10	The Risky Tasks in EUDAT	David Corney
16.20	Discussion	
16.45	Short break	
17.00	Community Presentations I	
18.30	End	
20.30	Dinner	

DAY2: 8TH MARCH 2012

09.00	SESSION 3 - ENABLING TECHNOLOGIES	CHAIR: MICHAEL LAUTENSCHLAGER
09.00	Distributed Authentication - will it work	Mark van de Sanden
09.10	PID Systems for Digital Objects	Ulrich Schwardmann
09.20	Replication Technologies	John Kennedy
09.30	Hosting Services and Staging Data	Johannes Reetz
09.40	Creating a Joint Semantic Domain	Peter Wittenburg
09.50	Other Technology and Operation Issues	M. van de Sanden/J. Reetz
10.00	Discussion	
10.30	Coffee Break	
11.00	SESSION 4 - EUDAT AND THE WAY FORWARD	CHAIR: PAVEL STRAŇÁK
11.15	Sustaining the infrastructure	Alison Kennedy
11.25	Training the new data scientist	Naghman Salman
11.35	Discussion	
12.00	Community Presentations II	
12.30	WRAP UP AND GENERAL DISCUSSION	K. KOSKI/ P. WITTENBURG
13.00	End	

Welcome to the 1st EUDAT Conference



5-8 November 2012, Barcelona

- A high level international event where EUDAT first results will be demonstrated
- A forum to discuss the future of EUDAT and data infrastructures
- 2nd EUDAT User Forum
- Training tutorials

EU-US collaboration

■ DAITF

- Data Access and Interoperability Task Force: to create a global interaction framework, pushing harmonization and standardisation with respect to the abstract data architecture and all its essential components.
- Data architecture workshop in Copenhagen (ICRI conference)

■ iCORDI (new project proposal)

- New 2M€ proposal submitted to the EC in November 2012 (results over eligibility for funding tba shortly)
- Project goal: to establish a **coordination platform** between Europe and the USA to discuss and improve the interoperability of today's and tomorrow's scientific data infrastructures of both continents.
 - fostering discussion between relevant stakeholders in the EU and US over concrete topics related to the interoperability of the data architectures and solutions based on a top-down approach;
 - overcoming the identified challenges and turning the areas of convergence into concrete specifications that can be immediately implemented on both continents by bringing data practitioners together in a bottom-up process;
 - demonstrating through concrete examples of collaboration what works and what are the remaining barriers and challenges to be tackled to achieve full interoperability.

EUDAT management team



Damien Lecarpentier, CSC,
Project Manager



Kimmo Koski, CSC
Project Coordinator



Peter Wittenburg, MPI-PL
Scientific Coordinator
Stakeholders requirements



Nagham Salman, BSC,
Dissemination



Alison Kennedy, EPCC
Sustainability



Mark van de Sanden, SARA
Services



Johannes Reetz, RZG
Operations



David Corney, STFC
Scalability